

DEPARTMENT OF ECONOMICS

WORKING PAPER SERIES

2004-17



McMASTER UNIVERSITY

Department of Economics
Kenneth Taylor Hall 426
1280 Main Street West
Hamilton, Ontario, Canada
L8S 4M4

<http://socserv.socsci.mcmaster.ca/~econ/>

**NONPARAMETRIC EXPENDITURE-BASED ESTIMATION OF
INCOME UNDER-REPORTING AND THE UNDERGROUND ECONOMY***

October 2004

Lindsay M. Tedds[†]

Department of Economics, McMaster University
426 Kenneth Taylor Hall, 1280 Main St. West, Hamilton, Ontario, L8S 4M4
Phone: 905-525-9140 ext. 24731, Fax: 905-521-8232
Email: teddsml@mcmaster.ca

ABSTRACT

There is considerable interest in measuring the underground economy using microeconomic data. One such method estimates income under-reporting by households by assuming a known, parametric form of the Engel curve and making the further parametric assumption that households under-report their income by a constant fraction, independent of income. This paper proposes a nonparametric approach which avoids functional form restrictions and enables the reporting function to vary across income levels and household characteristics. I illustrate by estimating the effect of the Canadian Goods and Services Tax on income under-reporting.

Keywords: Underground Economy, Income Under-reporting, Nonparametric Estimation, Engel Curve

JEL Classification: C14, D12, O17

* I would like to thank my supervisor Tom Crossley and Mike Veall and David Bjerk, all of McMaster University, David Giles and Herb Schuetze, both of the University of Victoria, and participants at seminars at McMaster University and the University of Victoria for their helpful comments and invaluable guidance. The author would also like to gratefully acknowledge the financial support from the Ontario Graduate Scholarship Program and the Social Sciences and Humanities Research Council (Doctoral Fellowship #752-2004-1096 and INE grant #501-2002-0107). I am solely responsible for any remaining errors and omissions.

[†] Lindsay M. Tedds is a graduate student in the Department of Economics at McMaster University. This paper stems from work that was undertaken by the author as part of the requirements for a Doctorate of Philosophy.

1. INTRODUCTION

There has been a recent resurgence in interest in measuring the underground economy and this interest has been stimulated predominantly by the perception that the underground economy is sizeable and growing. In broad terms, the phrase “underground economy” refers to output that is produced, and income that is generated, by agents who hide this fact from authorities. Knowledge of the size and structure of the underground economy is important for a number of reasons. First, because underground activities are unmeasured, they are not taken into account in the information-set that is used to assist economic policy-makers. Second, the underground economy effectively re-distributes both income and wealth in ways that are not necessarily consistent with the re-distributional goals of the taxation system. Third, the shortfall in income-reporting that is associated with underground activities leads to an erosion in the tax base and tax revenue, with subsequent implications for both public expenditure and taxation policies. Finally, enforcement activities are unlikely to be successful (and may have counter productive consequences) without detailed knowledge of the characteristics and types of activities of underground economy participants.

To date, research that seeks to measure the underground economy has predominately employed macro-methods.¹ These macroeconomic measures, however, have been criticized for not being consistent with modern economic models of consumer behaviour, employing flawed econometric techniques, producing unreliable estimates, and providing limited guidance to policy makers (Thomas 1999). In particular, the macro-

¹ Such methods include: the *Currency-Ratio Approach* (Gutmann 1977); the *Monetary-Transactions Method* (Feige 1979); *Tanzi's Approach* (Tanzi 1980); the *MIMIC model* (Frey and Weck-Hanneman 1984); and *National Accounts/Judgmental Methods*.

methods developed to date do not provide any information regarding the characteristics of those participating in the underground economy. In order to obtain this type of information, a method that utilizes microeconomic data is required.

One such approach, popularized by Pissarides and Weber (1989) and modified by Lyssiotou *et al.* (2004), utilizes household income and expenditure data to estimate the degree of income under-reporting (i.e. the amount by which household income should be scaled upwards to obtain true, or actual, income as opposed to reported income). The basic principle of this *Expenditure-based method* is that true household income can be imputed from reported household expenditures. The method is premised on variations of several key assumptions, namely: the reporting of expenditures on some items by all households is accurate; those who report zero self-employment income report income accurately while those who report nonzero self-employment income may under-report; and the marginal propensity to consume out of unreported income is equal to the marginal propensity to consume out of reported income. Actual, or true, self-employment income is then imputed by comparing the expenditure levels of households with positive self-employment income to the expenditure-income bundles of households with zero self-employment income and similar characteristics. In practice, the method is implemented by estimating reliable expenditure functions (i.e. Engel curves) for wage earners that are then inverted to estimate true income for the self-employed.

Previous studies have implemented the *Expenditure-based method* using highly parametric restrictions on: (1) an Engel curve (Pissarides and Weber 1989) or a system of Engel curves (Lyssiotou *et al.* 2004); and, (2) an income reporting function. These restrictions imply that households under-report their income by a constant fraction,

independent of income. There is no empirical evidence that supports this restriction and little, if anything, is actually known about the functional form of the reporting function. This paper considers an alternative way of implementing the *Expenditure-based method*. In particular, we relax the parametric restrictions and explore a nonparametric approach to the measurement of income under-reporting.

Specifically, we propose a two-step approach to estimating a variable-with-income reporting function, within the framework of the *Expenditure-based method*. The approach is essentially as follows. First, we estimate a nonparametric inverse food Engel curve for the sample of households that report zero self-employment income, to obtain an estimate of true income given (accurately) reported expenditures for every household in the sample (including those with self-employment income). Second, we estimate the nonparametric reporting function for self-employment income for households that report positive self-employment income. This approach improves on the implementation of the *Expenditure-based method* by minimizing the number of assumptions required for estimation. More particularly, the proposed framework avoids the usual functional form restrictions and enables the reporting function to vary across income levels and household characteristics.

The approach is illustrated by estimating the effect of the Canadian Goods and Services Tax (GST) on income under-reporting by married households with self-employment income. It is often argued that the implementation of this broadly based consumption tax increased the incentives and opportunities for tax evasion (Spiro 1993; Hill and Kabir 1996; Giles and Tedds 2002; and Tedds 2005). The empirical analysis uses the Canadian Family Expenditure Survey (FAMEX) which contains household level information about income and expenditures.

Overall, this refinement to the *Expenditure-based method* produces results which demonstrate that income under-reporting does vary across household income levels. In particular, income under-reporting is more prevalent among households at the lower end of the self-employment income distribution. Possible explanations of this finding are that households with more self-employment income may be more likely to be audited by the authorities, face higher utility costs if they are caught, or disproportionately benefit from legal tax avoidance (e.g. by exploiting various tax credits or loopholes). The results also provide evidence that that income under-reporting by married couples with self-employment income did increase following the implementation of the GST.

The remainder of this paper is organized as follows. Estimating income under-reporting from micro data is discussed in Section 2, which includes a brief overview of the literature and details regarding the non-parametric approach proposed by this paper. The application of the approach is described in section 3, including a description of the data, the results and a discussion. Section 4 concludes the paper.

2. *ESTIMATING INCOME UNDER-REPORTING FROM MICRO DATA*

2.1. *PREVIOUS APPROACHES*

In this section, attention is focused on two critical aspects of the empirical work in this paper with the view of placing the empirical strategy in context. These aspects concern: (1) functional form restrictions; and, (2) the treatment of permanent income.

2.1.1. FUNCTIONAL FORM RESTRICTIONS

A critical aspect of the empirical work in this area is the specification of the expenditure and reporting functions. The pioneering work in the development of the *Expenditure-based method* was conducted by Smith *et al.* (1986) and Pissarides and Weber

(1989).² First, they categorize households as either being self-employed or wage earning. Second, they specify a log-log (in expenditures and income) form for the expenditure equation (i.e. the constant elasticity Engel curve) that is used to estimate the parameter θ in the linear reporting function for self-employed households, defined as

$$y_{SE}^* = \theta y_{SE} \quad (1)$$

where y_{SE}^* represents true self-employment income, y_{SE} denotes reported self-employment income, and θ is assumed to be > 1 . This method of estimating income under-reporting consists of two steps. First, an expenditure function is estimated for wage earners. Second, the expenditure function is inverted to calculate θ , the amount by which reported self-employment income must be scaled up by in order to obtain true self-employment income.

Figure 1 provides a graphical representation of the approach. Constant-elasticity Engel curves for wage, or employee, and self-employed households are shown. A self-employed household reports expenditures, E^* , and income, Y , but the reported level of expenditures is actually consistent with true income, Y^* . The amount by which reported income must be scaled up to obtain true income is calculated by taking the ratio of the distance OY^*/OY which is equivalent to the parameter θ in equation (1) above. As the Engel curve for the self-employed is assumed to be parallel to that of wage earners, the distance is the same for every household (i.e. the reporting function is constant).

FIGURE 1 HERE

² The *Expenditure-based method* was developed following work conducted by Dilnot and Morris (1981) who calculated the difference between reported household income and expenditures and arbitrarily classify households as “black economy” households if expenditures exceeded income by at least 20 percent.

Lyssiotou *et al.* (2004) propose a systems approach to the *Expenditure-based method*. They specify a system of Engel curves of quadratic-in-income Working-Leser form. They assume that durable and nondurable goods are separable and base their demand system on nondurable goods only, namely: food, alcohol, fuel, clothing, personal goods/services, and leisure goods/services. Lyssiotou *et al.* (2004) maintain the specification of the linear reporting function given in equation (1) above but avoid classifying households as either wage earners or self-employed.³

The functional form for the Engel curve that is specified by Lyssiotou *et al.* (2004) raises two concerns. First, an implicit assumption of the *Expenditure-based method* is that the Engel curve(s) employed in the estimation must be monotonic in income. In reference to Figure 1, if this critical assumption is violated, then a unique value of true income associated with a particular level of expenditures may not exist. The quadratic-in-income Working-Leser form of the Engel curve specified by *et al.* (2004) is not necessarily consistent with this assumption, with particular goods, notably alcohol and clothing, known to violate this assumption (Banks *et al.* 1997). Second, the quadratic-in-income Working-Leser form of the Engel curve is not invertible over all values due to the presence of asymptotes. While the presence of asymptotes is not a concern under the structure imposed by Lyssiotou *et al.* (2004) - the system of Engel curves is not (implicitly) inverted over all

³ Lyssiotou *et al.* (2004) also allow for what they call “preference heterogeneity”. They note that income from self-employment may not be spent in the same way as income from other sources. In particular, it could be that households spend wage income, which is predictable, on necessities and the self-employment income, which is subject to under-reporting and is unpredictable, on luxuries. Equally, the self-employed could just have different preferences. Pissarides and Weber (1989) assumed homogenous preferences among all households. Lyssiotou *et al.* (2004) allow for preference heterogeneity in their estimated system of budget shares through the inclusion of the self-employment proportion of reported income, which can enter the system nonlinearly. The preference heterogeneity term(s), however, are identified only by functional form and are not identified in the nonparametric framework proposed in this paper.

data points - it underscores the likelihood that the estimates are influenced, in whole or in part, by the parametric restrictions.

More generally, this approach still assumes a parametric Engel curve, albeit one that is more general than that implied by earlier constant-elasticity assumption. Perhaps more importantly, this approach continues to assume that households under-report their income by a constant fraction, independent of income. In fact, little is known about the form of the reporting function and it is plausible that under-reporting will differ with income and household characteristics. This paper proposes a nonparametric approach which avoids functional form restrictions. The proposed method also works directly with an inverse Engel curve, avoiding problems associated with inversion, and continues with the tradition of the single equation approach (which implies that the restrictions required by consumer theory, notably homogeneity and adding-up restrictions, are not imposed). The single equation approach also allows the analysis to be restricted to a good for which the Engel curve is widely acknowledged to be monotonic.

2.1.2. PERMANENT VERSUS TRANSITORY INCOME

There is a general belief that households base expenditures on permanent rather than transitory income. This implies that households save when they have positive transitory income and dissave when they have negative transitory income. If the *Expenditure-based method* is implemented using transitory, or annual, income, this may lead to biased estimates of income under-reporting. Pissarides and Weber (1989) acknowledge that permanent income is the measure of income that influences consumption decisions but stop short of requiring their expenditure function to conform exactly to the permanent income hypothesis, perhaps because the data set used in their analysis (1982 British Family

Expenditure Survey) did not contain information regarding household savings behaviour. They indicate that "...for given permanent income, the measured income of the self-employed may be more variable than the measured income of employees in employment. If this is correct, our measure of income under-reporting by the self-employed will have to be adjusted accordingly." (Pissarides and Weber 1989, 20) Empirically, they implement this assumption by treating reported income as endogenous and then using instrumental estimation, which "...enables an independent estimate of the residual variance of reported income for each group which we exploit in the calculation of income under-reporting." (Pissarides and Weber 1989, 22)

Whether Pissarides and Weber's (1989) 2SLS approach is preferred to OLS depends on the quality of the instruments. Data sets that contain information on household expenditures and income may not contain relevant instrumental variables required for this analysis. Further, the approach requires the researcher to make additional, and somewhat arbitrary, assumptions which restrict the analysis. As a result, an alternative approach which addresses the issue of permanent income is desirable. This paper explores such an alternative.

2.2. A NONPARAMETRIC APPROACH

As outlined above, to date, the *Expenditure-based method* has been implemented by estimating Engel curves which are implicitly or explicitly inverted to obtain an average estimate of income under-reporting. A more direct approach to estimating income under-reporting is to utilize an inverse Engel curve (i.e. with income taking on the role of the dependent variable) and nonparametric methods. Within the framework of the *Expenditure-based method*, I propose a two-step approach to estimating a variable-with-

income reporting function that responds to the concerns raised in the previous section. The first step nonparametrically estimates an inverse Engel curve, which can be consistently estimated for households that report zero self-employment income, to obtain true income for all households. The second step nonparametrically estimates the reporting function for households with positive self-employment income.

The use of nonparametric methods⁴ here has three advantages. First, it enables the reporting function to vary across income levels and household characteristics. Second, it avoids functional form restrictions on the Engel curve. Third, within this framework I am also able to test the null hypothesis that the reporting function is linear, as has been assumed in the previous literature.

To achieve estimation, some initial assumptions are required. I maintain the three fundamental assumptions of Pissarides and Weber (1989) but avoid classifying households as either self-employed or not, following Lyssiotou *et al.* (2004). First, food expenditures are used in the analysis and it is assumed that the reporting of food expenditures by all households is accurate.⁵ Second, only self-employed income can be under-reported.⁶

⁴ Semiparametric estimation was not pursued as Blundell *et al.* (1998) have found that "...the additive structure between demographic composition and income that underlies the partially linear semiparametric model implies strong and unreasonable restrictions on behaviour." (Blundell *et al.* 1998, 461). The nonparametric estimation strategy proposed here cannot be implemented if income and demographic terms enter non-additively. Instead, estimation is conducted separately on identified homogenous sub-populations (i.e. single households, married couples, etc.).

⁵ The arguments for using food as opposed to any other commodity or group of commodities are that: there is no social stigma associated with food consumption which could cause expenditures to be reported inaccurately (counter examples would include tobacco and alcohol); tastes for food are more likely to be uniform across employment groups and over time; it is very difficult for a household to postpone food consumption; most food purchases cannot be included as a business expense; the food Engel curve is widely acknowledged to be monotonic; and, food expenditures are more likely to be reported accurately by households participating in the underground economy since individual expenditures on food are small and are unlikely to rouse suspicion.

⁶ Taxes for most sources of income, particularly wage and salary income, and various "payroll" taxes are "pay as you earn". That is, income and payroll taxes are withheld at source from these payments to individuals.

Third, the marginal propensity to consume out of unreported income is constrained to be equal to the marginal propensity to consume out of reported income.⁷

The approach in this paper also considers the issue of permanent versus transitory income. As indicated above, there is a general belief that households base expenditures on permanent rather than transitory income. This implies that households save (dissave) when transitory income is greater (less) than permanent income. If transitory, rather than permanent, income is used when estimating income under-reporting then it is likely that (dis)savings activity is being confused with misreporting. For example, a person with temporarily low income will dissave in order to maintain consumption but, if savings are ignored in the analysis, this behaviour will be indicative of income under-reporting. The result will be a biased estimate of income under-reporting. If, however, one can observe whether a household is saving or dissaving, then this information, along with reported income, can be used to approximate permanent income and used in the estimation of income under-reporting, which obviates the need for the 2SLS approach followed by Pissarides and Weber (1989) described in the previous section.

Self-employment income, on the other hand, is reported and taxed at year end (though many self-employed are required to make estimated tax payments during the year in order to ensure that they meet their tax obligation in a timely manner) by the individual who earned the income and there is no third party who also reports this income. That is, there is no check and balance within the system to ensure that the individual is accurately reporting their self-employment income. As a result, there is an opportunity for some self-employment income to be under-reported. That said, the assumption that only self-employment income is under-reported is likely not entirely accurate. For example, employers can pay their employees in whole or partially in cash as a way to evade income and payroll taxes. The extent that this assumption is not valid will lead to the resulting estimate of the degree of under-reporting to be biased toward zero.

⁷ The reader should be made aware that this assumption may not be accurate. It may not be true that the marginal propensity to consume out of unreported income is equal to the marginal propensity to consume out of reported income. Households may use all unreported income to boost expenditures. Alternatively, households may use unreported income to boost savings, though the inclusion of the net change in assets and liabilities in the analysis will likely account for this behaviour.

The estimation strategy follows is as follows. The object of interest is true household self-employment income, y_{SEh}^* , which is assumed to be a function of reported household self-employment income, y_{SEh} , plus a white noise disturbance term:

$$E(y_{SEh}^* | y_{SEh}, d = 1] = f(y_{SEh}) + \xi_h \quad (2)$$

where h denotes an individual household and d is a dummy variable that takes a value of 1 if the household reports any self-employment income.

The first stage of the procedure is to nonparametrically estimate an inverse Engel curve to obtain true (permanent) income given (accurately) reported expenditures. The inverse Engel curve expresses income, in this case permanent income for reasons discussed above, as a function of expenditures. For this exercise, the nonparametric representative of the inverse Engel curve is given by:

$$y_{TOTALh}^p = h(x_h) + v_h \quad (3a)$$

where x_h represents household reported (and assumed to be true) food expenditures, v_h is a white noise disturbance term, and y_{TOTALh}^p represents true (reported plus unreported) total permanent household income, defined as

$$y_{TOTALh}^p \equiv y_{SEh}^* + y_{OTHh} - \Delta A_h \quad (3b)$$

y_{OTHh} refers to household reported (and assumed to be accurately reported) other income and ΔA_h indicates household net change in assets and liabilities (a households that has positive (negative) transitory income will save (dissave) the additional money and $\Delta A_h > 0$ (< 0)).

By assumption, x_h is accurately observed for all households but y_{TOTALh}^p is only accurately observed for those households that have zero self-employment income ($y_{SEh}^* = y_{SE} = 0$). This implies that $h(x_h)$ can be consistently estimated for households that report zero self-employment income. The fitted values from the first stage regression, $\hat{h}(x)$, for households that report zero self-employment income are used to obtain an accurate estimate of total permanent income for households with positive self-employment income based on food expenditures. As a result, consistent estimates of total permanent household income, $\hat{h}(x)$, are obtained for every household.

As indicated in equation (3b) above, total permanent household income is comprised of three elements, namely the household's: true self-employment income (y_{SEh}^*), reported other income (y_{OTHh}), and net change in assets and liabilities (ΔA_h). If y_{OTHh} is subtracted from, and ΔA_h is added to, the estimate of total permanent household income obtained in the first step, $\hat{h}(x)$, one obtains an estimate of true self-employment income, y_{SEh}^* , for those households that report positive self-employment income. That is, y_{SEh}^* can be calculated as follows:

$$y_{SEh}^* = \hat{h}(x_h) - y_{OTHh} + \Delta A_h \quad (3c)$$

which is used in the second step of this approach.

The second step estimates the nonparametric form of the reporting function, the parametric form of which is given by equation (1), for those households which report positive self-employment income ($y_{SE} > 0$). The nonparametric form of the reporting function is given by:

$$y_{SEh}^* = f(y_{SEh}) + \zeta_h \quad . \quad (4)$$

The amount of self-employment income that is unreported by each households is calculated as the predicted value of true self-employment income $\hat{f}(y_{SE})$ minus reported self-employment income y_{SE} . Total unreported income is found by summing over households with positive reported self-employment income.

2.3. TESTING LINEARITY OF THE REPORTING FUNCTION

As indicated above, previous studies assumed that the reporting function took the form denoted in equation (1), where θ is assumed to be > 1 . The nonparametric approach outlined above provides an opportunity to test the null hypothesis that the reporting function takes the linear form specified by equation (1) versus the alternative that the reporting function takes the nonparametric specification specified by equation (4).

To implement this test, I utilize a testing method described by Yatchew (1998). The test statistic is given by

$$V = \frac{T^{1/2}(s_{res}^2 - s_{diff}^2)}{s_{diff}^2} \sim N(0,1) \quad (6)$$

where

$$s_{diff}^2 = \frac{1}{2T} \sum (y_{SE,t}^* - y_{SE,t-1}^*)^2 \quad (7)$$

$$s_{res}^2 = \frac{1}{T} \sum (y_{SE,t}^* - \hat{\theta} y_{SE,t})^2 \quad (8)$$

and T is the number of households.

The testing procedure is as follows. First, the data are reordered such that

$y_{SE,1} \leq \dots \leq y_{SE,T}$. Second, s_{diff}^2 is calculated. Third, the restricted regression given by

equation (1) is performed to obtain $y_{SE,t}^* - \hat{\theta} y_{SE,t}$. Fourth, s_{res}^2 is calculated. Finally, the test statistic, V , is calculated and a one-sided test is conducted, comparing the value of the test statistic to a critical value from a standard normal distribution.

2.4. TESTING THE SIGNIFICANCE OF THE CHANGE IN ASSET TERM

It is also possible to test the significance of ΔA , the change in assets term, in equation (3) by employing the differencing method discussed in Yatchew (1998, 2003). To do so, note that equation (3) can be rewritten as

$$y_h^a = h(x_h) + \beta \Delta A_h + v_h \quad (9)$$

where y_h^a represents a households annual income (where $y_h^a = y_{SEh}^* + y_{OTHh}$). Equation (9) is a partially linear model in ΔA_h . In equation (3) above, β was assumed to be equal to 1.

In order to test if $\beta=0$ or, alternatively, if $\beta=1$, the data must first be sorted such that $x_1 \leq \dots \leq x_T$. The variables y_h^a and ΔA_h are then differenced (which, in heuristic terms, “removes” the direct effect $h(x)$ of the nonparametric variables x that occurs through ΔA_h). We can then apply the OLS estimator to the differenced data such that:

$$\hat{\beta}_{diff} = \frac{\sum (y_h^a - y_{h-1}^a)(\Delta A_h - \Delta A_{h-1})}{(\Delta A_h - \Delta A_{h-1})^2} \quad (11)$$

The process of differencing the data, however, creates autocorrelation in the error term. Yatchew (2003) notes that the correction is simple if homoskedasticity is assumed: the standard errors simply need to be multiplied by the square root of 1.5. Following this correction, standard inference techniques can be employed.

3. APPLICATION

The nonparametric application of the *Expenditure-based method* outlined above is illustrated here by estimating the effect of the Canadian Goods and Services Tax on income under-reporting.

3.1. DATA

The data used in this paper come from the public use Canadian Family Expenditure Surveys (FAMEX), which were conducted at irregular intervals between 1969 and 1996.⁸ The FAMEX is a cross-sectional household recall survey that is intended to be representative of all persons living in private households in the 10 provinces in Canada.⁹ (Additional information on the collection of the FAMEX is provided in Appendix A.) The sample for this analysis is limited to married couples (without children) and it is assumed that the household unit acts as a single decision maker regarding expenditure and income reporting. (A short discussion concerning the unit of analysis is provided in Appendix B.)

The sample is further restricted to households: where the head and spouse are of working age (25-64 years of age); which constitute one economic family; that have positive food expenditures; and for which the head occupation is known and is not working in the primary occupation category. (This last restriction will exclude farm households, which are likely to have much different expenditure patterns on food than those in other occupations.) Households whose annual gross income was either in the top or bottom 1 percent of the income distribution were excluded from the analysis. In addition, households whose

⁸ In 1997, the Survey of Household spending (SHS) replaced the FAMEX and has been conducted annually since. The SHS, however, does not provide detailed information regarding the sources of household income so this data cannot be used for this analysis.

⁹ Households in the Territories are also surveyed but their data is not included in all the public use files.

permanent gross income¹⁰ was either in the top or bottom 1 percent of the income distribution were also excluded from the analysis. These last two exclusions are intended to avoid households with negative income and extreme positive income in both steps of the method described in section 2.2 above. Finally, households with negative self-employment income are also excluded from the analysis.

The implementation of the GST in 1991 has probably received the most “credit” for increasing the size and growth of the Canadian underground economy in recent times. It is often argued that the switch from the more narrowly based federal manufacturers’ sales tax to this broadly based consumption tax may have increased the incentives and opportunities for tax evasion (Spiro 1993; Hill and Kabir 1996; Giles and Tedds 2002; and Tedds 2005). The framework described in the previous section provides an opportunity to explore this hypothesis.

To conduct the analysis, results from using FAMEX data for the years 1982 and 1986 will be compared to those obtained using data for the years 1992 and 1996. Pooling the data in this way attempts to ensure that there are sufficient observations included in each stage of the analysis. Each pooled sample contains one year during which the economy was sluggish (1982 and 1992) and one year in which the economy was in a growth period (1986 and 1996). The implicit restriction made by pooling the data in this way is that the marginal propensity to consume food is the same for each of the two years contained in each of the pooled samples. Two additional households in the pooled 1982/1986 sample were excluded from the analysis as well as one additional household in

¹⁰ This is the dependent variable in the first stage regression and is defined as gross income less change in assets.

the pooled 1992/1996 sample. These households had self-employment income that exceeded average self-employment income by a factor of almost six. As there were no other observations within their vicinity it was not possible to obtain nonparametric estimates at these points by using any reasonable bandwidth. Pooling, along with the restrictions noted here and above, left a total of 1,907 households in the 1982 and 1986 pooled sample, of which 303 are self-employed and a total of 1,840 households in the 1992 and 1996 pooled sample, of which 369 are self employed. The increase in the ratio of self-employed households to non self-employed households between the two samples is not unexpected given that the Canadian self-employment rate rose from 13 percent in 1979 to 18 percent by 1997 (Picot *et al.* 1998).

Expenditures are converted to real 1996 dollars using the food price index developed by Browning and Thomas (1999). Food expenditures, which includes expenditure on food consumed at home and in restaurants, are used in estimating equation (3).¹¹ Income terms and the change in asset term are converted to real 1996 dollars using a general price index that was developed by employing the same methodology as Browning and Thomas (1999). All income terms are inclusive of income taxes as net income by source is not available in the FAMEX.¹²

Table 1 provides some summary statistics of the data. The top half of the table presents statistics for households with zero self-employment income while the bottom half of the table presents statistics for households with positive self-employment income. The left column shows statistics for the 1982/1986 pooled sample and the right column for

¹¹ Similar estimates to those reported in section 3.2 were obtained when food expenditures were restricted to include only expenditures on food consumed at home.

¹² Pissarides and Weber use net income in their analysis.

1992/1996. The two household groups report comparable average incomes, changes in assets, and expenditures on food in each of the two samples, but self-employed households have greater variability in their assets in the 1982/1986 sample.

TABLE 1 HERE

3.2. RESULTS

Nonparametric estimation of equations (3) and (4) is achieved by employing the locally-weighted least-squares procedure, using the Gaussian weighting function and adaptive bandwidth¹³. Equation (3), the inverse Engel curve, is estimated at every point in the data but assigning a weight of zero to households with positive self-employment income in the estimation process. The reporting function given by equation (4) is estimated only for those households which report positive self-employment income ($y_{SE} > 0$).

As outlined in section 2.4, it is possible to test the significance of the ΔA term in equation (3a). The results of this test are outlined in Table 2. As before, the results for 1982/1986 are in the column on the right and 1992/1996 are presented in the left-hand column. The parameter estimates for β_{diff} , noted in the first row, are very close to unity in value. In both cases, we reject the null hypothesis that $\beta_{diff} = 0$ with p-values of essentially zero, as is noted in the second row of the table. The results for testing the null hypothesis that $\beta_{diff} = 1$ are shown in the third row. For the 1982/1986 pooled data set, we would fail to reject the null hypothesis that $\beta_{diff} = 1$ at the 1% or 5% significance levels but would reject it at a 10% significant level. For the 1992/1996 pooled data set, we fail to reject the null hypothesis that $\beta_{diff} = 1$ at any conventional significance level. Given the

¹³ The initializing bandwidth was selected by cross-validation (Härdle and Marron 1990).

test results and the fact that the estimates for β_{diff} are economically no different from unity, we conclude that the ΔA term should be included in the analysis as outlined in section 2.2 and proceed accordingly.

TABLE 2 HERE

Figure 2 presents graphs of the inverse food Engel curve, estimated from equation (3a). Recall from section 2.2 that equation (3a) can be consistently estimated on the sample of households that report zero self-employment income and provides an estimate of true household income for all households. The graph on the left is for the 1982/1986 pooled sample while the graph on the right is for 1992/1996. Reported food expenditure is plotted on the horizontal axis and gross household income less changes in assets is plotted on the vertical axis. In both case, the inverse food Engel curve appears linear over most food expenditures but takes on some curvature at higher levels of food expenditures, notably where the data becomes sparse.¹⁴

FIGURE 2 HERE

Figure 3 presents graphs of the nonparametrically estimated amounts of true household self-employment income, $\hat{f}(y_{SE})$, that were obtained from equation (4). Again, the graph on the left is for the 1982/1986 pooled sample while the graph on the right is for 1992/1996. Estimated true self-employment income is plotted on the vertical axis and reported self-employment income is plotted on the horizontal axis. Both axes use the log

¹⁴ The inverse Engel curves obtained from equation (3) but without the change in asset term (where gross income is the dependent variable), are similar in shape to those shown in Figure 2 but are shifted vertically. Ignoring the change in asset term, therefore, leads to inflated estimates of true gross income given reported food expenditures.

scale. Also shown are 90% bootstrapped confidence intervals obtained using the “wild” bootstrap procedure (Wu 1986) which allows for heteroskedastic errors. The 45 degree line in the figures shows reported self-employment income. When the plot of estimated true self-employment income is above the 45 degree line, a household is under-reporting their self-employment income. Each graph also presents three vertical lines which represent the 10th, 50th, and 90th percentiles of the data. This information is presented to provide the reader with detail regarding the density of the data and its relation to the estimation of the reporting function.

The graphs in Figure 3 show that the reporting function appears to be nonlinear. For the 1982/1986 pooled sample, estimated true self-employment income is above reported self-employment income for households with less than almost \$40,000 in reported self-employment income but under-reporting decreases as reported self-employment income approaches approximately \$40,000. For the 1992/1996 pooled sample, estimated true self-employment income is above reported self-employment income for households with less than just over \$40,000 in reported self-employment income but under-reporting decreases as reported self-employment income increases beyond approximately \$40,000. Beyond the approximate \$40,000 threshold amount in both samples, the results indicate that households over-report self-employment income. The reader should note, however, that the estimated number of married households that over-report is small in percentage terms. Moreover, in both pooled samples, the data are sparse beyond \$40,000. In the 1982/1986 pooled sample, the 90th percentile occurs at approximately \$46,800 (\$55,000 in the 1992/1996 pooled sample). In both cases, the 90th percentile occurs in the vicinity of

where estimated true self-employment income falls below reported self-employment income.

FIGURE 3 HERE

As mentioned in section 2.3 above, it is possible to test whether or not the reporting function, equation (4), is linear, as assumed previously in the literature. Table 3 summarizes the results of the test of null hypothesis that the reporting function takes the form of equation (1) against the alternative that the reporting function takes the nonparametric specification of equation (4). The results for the 1982/1986 pooled data set are noted in the first column. The value of the test statistic is noted in the first row and the associated p-value is reported in the second row. We obtain a value for the test statistic of 1.306 with an associated p-value of 0.096, hence, we reject the null hypothesis

$H_0 : y_{SE}^* = \theta y_{SE}$ in favour of the alternative $H_a : y_{SE}^* = f(y_{SE})$ at the 10% significance level. For the 1992/1996 pooled data set, the results of which are reported in the column on the left, we obtain a value for the test statistic of 2.863, noted in the first row, with an associated p-value of essentially zero, shown in the second row. We therefore reject the null hypothesis at all of the usual significance levels.

While not shown here, the second stage results obtained when the change in assets term is excluded from the analysis differ from those obtained when the change in asset term is included in the analysis. The reporting function for households with self-employment income less than \$40,000 are similar in both cases, with those obtained ignoring the asset term being only slightly higher than those shown in Figure 3. The dramatic difference occurs at self-employment incomes greater than \$40,000. The reporting function obtained

ignoring the asset term for these higher income levels flattens out immediately while the reporting function obtained including the asset term continues its upward trend, as shown in Figure 3. This implies that using annual income as opposed to permanent income in the analysis will result in a slight overstatement of income under-reporting for households with lower amounts of reported self-employment income and a sharp understatement of under-reporting by households with higher amount of reported self-employment income.

TABLE 3 HERE

Table 4 reports household population estimates of income under-reporting by the Canadian self-employed for 1982/1986, presented in the column on the left, and 1992/1996 in the column on the right. The total amount of income under-reporting is found by subtracting reported self-employment income from estimated household true self-employment income and summing up over households. Survey weights are used in the final step to obtain population estimates, noted in the first row of the table. The first row of table 4 shows the results for total income under-reporting. Total income under-reporting almost doubled between the 1980's and the 1990's, amounting to just over \$0.619B in the 1982/1986 pooled sample and increasing to approximately \$1.198B in the 1992/1986 pooled sample. The associated 90% bootstrapped confidence intervals are noted in the parenthesis. There are two things to note with respect to the reported confidence intervals. First, for both samples, the confidence intervals indicate that total income under-reporting was statistically significantly greater than zero. Second, as the confidence intervals overlap this suggests that total income under-reporting in 1992/1996 was not statistically significantly different from total income under-reporting in 1982/1986. Further statistical tests confirm that the difference is not statistically significant. The difference is, however,

economically significant. That is, an increase in total income under-reporting of \$0.579B by married households that report positive self-employment income between the two pooled samples is of enough magnitude to have an economic impact.

As the number of self-employed households increased between these two pooled sample, as shown in the second row of table 4, it could be that the increase in total income under-reporting was simply due to the increase in self-employed households over the sample period, rather than due to the implementation of the GST. In order to determine if there was a change in the amount of income under-reporting per household, average per household income under-reporting is calculated.¹⁵ Despite the fact that the number of self-employed households increased between these two pooled samples, there was an increase in the average amount of self-employment income that is unreported. Income under-reporting per married household, presented in the third row, amounted to \$2,462.70 in the 1982/1986 pooled sample and \$3,015.71 in the 1992/1996 pooled sample. The 90% bootstrapped confidence intervals for these per household amounts are presented in the final row of the table. Again, for both samples, the confidence intervals indicate that average income under-reporting is statistically significant. Further, while the difference is not statistically significant, as with total income under-reporting, a strong argument can be made that an increase in average income under-reporting of \$553 by married households that report positive self-employment income between the two pooled samples is economically significant.

3.3. LIMITATIONS

¹⁵ Average income under-reporting per married household with positive reported self-employment income is calculated by dividing total income under-reporting, reported in the first line of table 4, by the population size, also reported in table 4.

The results presented above call into question many of the assumptions made in the parametric approach of the *Expenditure-based method*. That said, some caution needs to be exercised in interpreting these specific results as the reliability of the estimate depends on the quality of the data. In particular, in using survey data, we are limited to studying only those households that have taken part in the survey. Households that are heavily involved in underground activity, particularly those households that are involved in illegal activity (such as drug trafficking, human smuggling and prostitution, for example), are unlikely to participate in the survey or may elect to modify their reported amount of expenditures to ensure they are not perceived to be living beyond their means.

Caution also needs to be exercised in interpreting and comparing the results presented here to those obtained by alternate methods. The results presented here, income under-reporting by married households with self-employment income, should not be interpreted as representing a measure of the total underground economy. Households with self-employment income but with different demographic characteristics (e.g. households with children, single person households etc.) may engage in income under-reporting at different rates than married households. Additionally, income under-reporting by the self-employed represents only a portion of underground activity. Finally, the method presented in this paper estimates income that is not reported to tax authorities, which is quite distinct from measuring production or income that is missed by the statistical offices when they calculate the value of the national product. Many methods employed in estimating underground activity use the latter calculation. Giles and Tedds (2002), updated by Tedds (2005), provide a summary of the available Canadian estimates of underground activity,

arranged according to methodology and calculation employed should the reader which to make their own comparisons.

4. CONCLUSION

This paper proposes a nonparametric approach for estimating income under-reporting by households with self-employment income. The use of nonparametric methods is shown to have several advantages over previous parametric approaches. First, it enables the reporting function to vary across income levels and household characteristics. Second, we are able to test, and find evidence against, the previously held hypothesis that the reporting function takes the linear form. Third, the framework allowed for an alternative approach to addressing the issue of permanent income. A further advantage of this method is the ease in which population estimates can be generated. In particular, we are able to obtain the total amount of unreported income in the population directly, whereas previous studies could only extrapolate this information by using national account data. Overall, the approach outlined in this paper calls into question many of the assumptions made in the parametric applications of the *Expenditure-based method*.

The approach is illustrated by estimating the effect of the Canadian Goods and Services Tax on income under-reporting by married households with self-employment income. The results indicate that income under-reporting by married households with self-employment income increased following the implementation of the GST. The results indicate that income underreporting increased, in real (1996) dollar terms, from \$2,462.70 per household in the 1980's to \$3,015.71 per household in the 1990's following the implementation of the GST. The paper does note that caution needs to be exercised in interpreting these specific results as the reliability of the estimate depends on the quality of

the data and on the various assumptions made. Evidence is provided which support the notion that the obtained estimates of income under-reporting reported in this paper are lower bound estimates.

The analysis presented in this paper indicates that further work is required in refining this method such that it is more consistent with available data and knowledge concerning participation in the underground economy. One possible refinement would be to conduct the analysis using reported taxes in place of permanent income. Implementing this refinement would require households to be classified as either self-employment or not as was done by Pissarides and Weber (1989). Additionally, with respect to the Canadian estimates presented in this paper, it is possible to conduct the analysis using a different data set. As indicated in section 3.3.1, the data for the FAMEX is collected in March/April of a given year, but covers expenditures for the previous year and the data collectors make attempts to ensure that total expenditures are roughly equal to total income. The Family Food Expenditure Survey (FOODEX), on the other hand, uses the diary system to collect its data. As this could result in more accurate expenditure data, it would be interesting to compare the results obtained from the FAMEX data with those obtained using the FOODEX data.

Figure 1: Income Under-reporting in the Single Equation Expenditure-Based Method

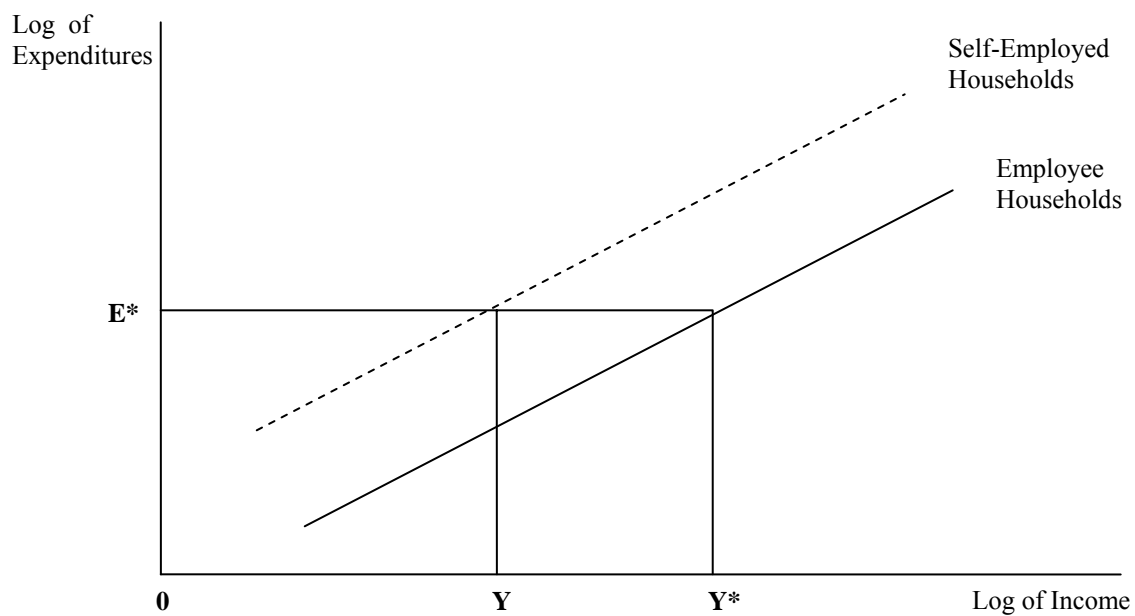


Table 1: Data Summary^a

	1982 & 1986 Pooled FAMEX				1992 & 1996 Pooled FAMEX			
Households with zero self-employment income								
	Sample Size=1604				Sample Size=1471			
	Mean	Std dev	Min	Max	Mean	Std dev	Min	Max
Gross Income (\$)	60,343	25,541	14,877	159,661	64,741	28,103	16,131	183,370
Self-Employment Income (\$)	0	0	0	0	0	0	0	0
All Other Income (\$)	60,343	25,541	14,877	159,661	64,741	28,103	16,131	183,370
Expenditures on Food (\$)	6,660	2,552	926	18,672	6,103	2,551	495	19,678
Net change in assets and liabilities (\$)	6,046	12,275	-32,892	83,869	6,086	13,732	-56,120	88,171
Households with positive self-employment income								
	Sample Size=303				Sample Size=369			
	Mean	Std dev	Min	Max	Mean	Std dev	Min	Max
Gross Income (\$)	55,808	27,372	15,041	137,853	60,545	29,283	15,244	184,000
Self-Employment Income (\$)	19,612	20,080	56	94,436	22,019	24,453	1	126,356
All Other Income (\$)	36,196	28,216	0	132,936	38,527	27,139	0	132,674
Expenditures on Food (\$)	6,365	2,727	1,400	17,536	6,061	2,681	1,489	18,359
Net change in assets and liabilities (\$)	5,785	16,020	-46,205	76,870	5,582	13,907	-42,619	61,211

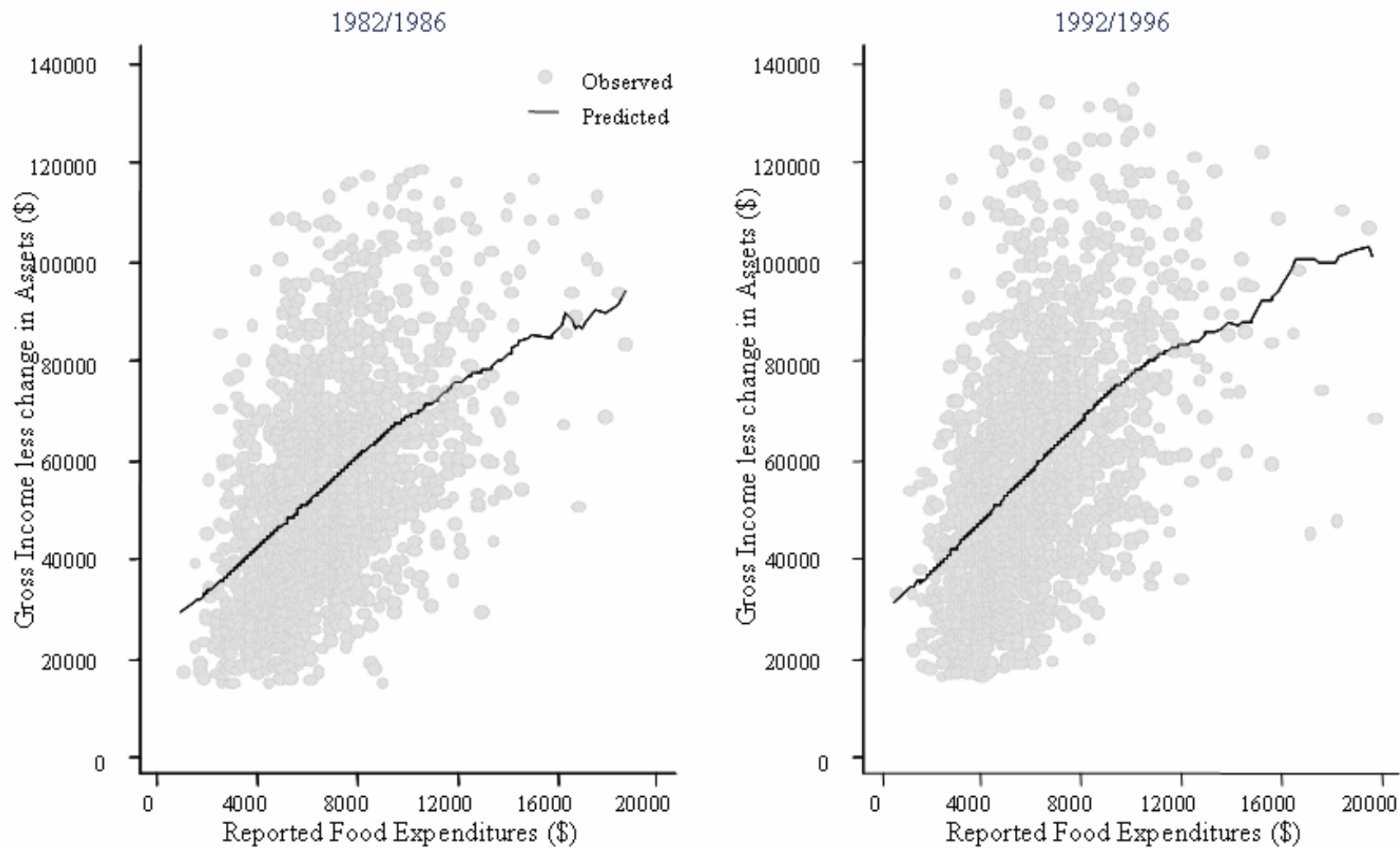
^a Amounts are in real (1996) Canadian dollars and are rounded to the nearest dollar.

Table 2: Testing the Significance of ΔA

	1982 & 1986 Pooled FAMEX	1992 & 1996 Pooled FAMEX
$\hat{\beta}_{diff}$ (<i>s.e.</i>)	1.053 (0.0408) ^a	1.047 (0.0440) ^a
Test: $H_0 : \beta_{diff} = 0$ vs. $H_a : \beta_{diff} \neq 0$	p-value=0.000	p-value=0.000
Test: $H_0 : \beta_{diff} = 1$ vs. $H_a : \beta_{diff} \neq 1$	p-value=0.098	p-value=0.143

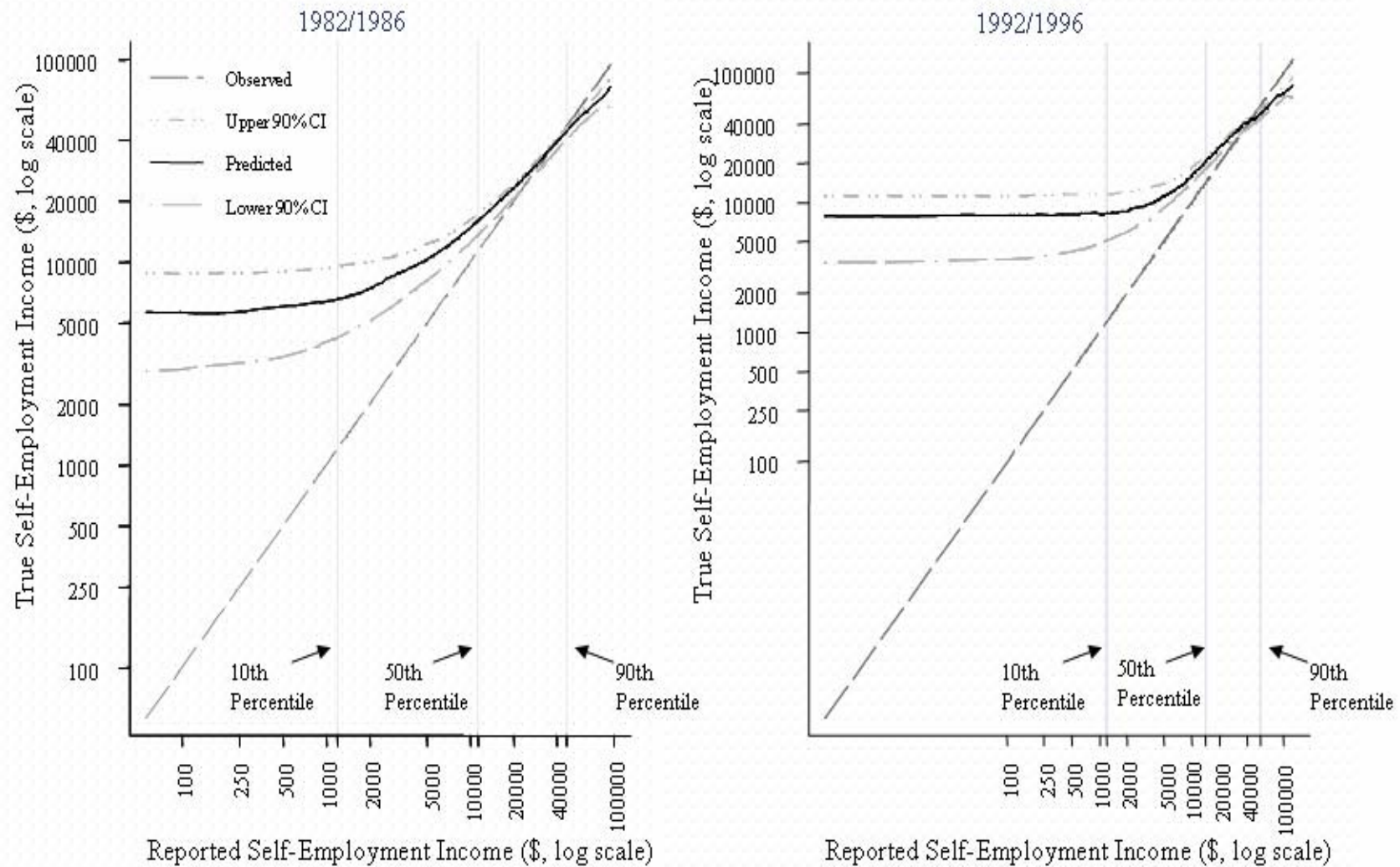
^a Standard errors corrected for autocorrelation as discussed in section 2.4.

Figure 2: Inverse Food Engel Curves



Source: FAMEX, Real 1996 \$, Married Couples

Figure 3: Estimated True Self-Employment Income (log scale)



Source: FAMEX, Real 1996 \$, Married Couples

Table 3: Testing Linearity of the Reporting Function

Test: $H_0 : y_{SE}^* = \theta y_{SE}$ vs. $H_a : y_{SE}^* = f(y_{SE})$

	1982 & 1986 Pooled FAMEX	1992 & 1996 Pooled FAMEX
V	1.306	2.863
p-value	0.096	0.002

Table 4: Estimates of Income Under-Reporting^a

	1982 & 1986 Pooled FAMEX	1992 & 1996 Pooled FAMEX
Total amount of Income Under-Reporting^b (90% Bootstrapped Confidence Interval)	\$0.619 Billion (\$0.116B; \$1.086B)	\$1,198 Billion (\$0.612B; \$2.358B)
Population Size^b	251,386	397,189
Income Under-Reporting Per Married Household^b (90% Bootstrapped Confidence Interval)	\$2,463 (\$428; \$4,278)	\$3,016 (\$1,542; \$5,936)

^a Amounts are in real (1996) Canadian dollars and are rounded to the nearest dollar

^b Calculated for married households that report positive self-employment income using the survey weights provided in the FAMEX by Statistics Canada to obtain population amounts.

BIBLIOGRAPHY

Banks, James, Richard Blundell and Arthur Lewbel (1997). "Quadratic Engel Curves and Consumer Demand", *Review of Economics and Statistics*, vol. 4, pp. 527-538.

Blundell, Richard, Alan Duncan and Krishna Pendakur (1998). "Semiparametric Estimation and Consumer Demand", *Journal of Applied Econometrics*, vol. 13, pp. 435-461.

Browning, Martin and Irene Thomas (1999). "Prices for the FAMEX: Methods and Sources", Department of Economics, McMaster University, Working Paper.

Cagan, Phillip (1958). "The Demand for Currency Relative for the Total Money Supply", *Journal of Political Economy*, vol. 66, pp. 302-328.

Dilnot, A.W. and C.N. Morris (1981). "What do We Know About the Black Economy", *Fiscal Studies*, vol. 2, pp. 58-73.

Feige, Edgar L (1979). "How Big is the Irregular Economy?", *Challenge*, vol. 22, pp. 5-13.

Frey, B.S. and H. Weck-Hanneman (1984). "The Hidden Economy as an Unobserved Variable", *European Economic Review*, vol. 26, pp. 33-53.

Giles, David E.A. and Lindsay M. Tedds (2002) Taxes and the Canadian Hidden Economy Canada Tax Foundation.

Gutmann, Peter M (1977). "The Subterranean Economy", *Financial Analysts Journal*, vol. 34, pp. 24-27.

Härdle, Wolfgang and James S. Marron (1990). "Semiparametric comparison of Regression Curves", *Annals of Statistics*, vol. 18, pp. 63-89.

Hill, Roderick and Muhammed Kabir (1996). "Tax Rates, the Tax Mix, and the Growth of the Underground Economy in Canada: What Can We Infer", *Canadian Tax Journal*, vol. 44, pp. 1552-1583.

Lyssiotou, Panayiota, Panos Pashardes and Thanasis Stengos (2004). "Estimates of the Black Economy Based on Consumer Demand Approaches," *Economic Journal*, vol. 114, pp. 622-639.

Picot, G., Marilyn Manswer and Zhengxi Lin (1998). "The Role of Self-Employment in Job Creation in Canada and the United States", US Bureau of Labor Statistics and Statistics Canada, Working Paper.

Pissarides, Christopher A., and Guglielmo Weber (1989). "An Expenditure-based Estimate of Britain's Black Economy," *Journal of Public Economics*, vol. 39, pp. 17-32.

Smith, Stephen, Christopher A. Pissarides and Guglielmo Weber (1986). "Evidence from Survey Discrepancies", in Britain's Shadow Economy, edited by Stephen Smith, Clarendon Press, pp. .

Spiro, Peter S. (1993). "Evidence of a Post-GST Increase in the Underground Economy", *Canadian Tax Journal*, vol. 41, pp. 247-258.

Tanzi, Vito (1980). "The Underground Economy in the United States: Estimates and Implications", *Banco Nazionale del Lavoro*, vol. 135, pp. 427-453.

Tedds, Lindsay M (2005). "The Underground Economy in Canada", in Size, Causes and Consequences of the Underground Economy, edited by Chris Bajada and Friedrich Schneider, Ashgate Publishing (forthcoming).

Thomas, Jim (1999). "Quantifying the Black Economy: Measurement without Theory Yet Again", *Economic Journal*, vol. 109, pp. F381-F337.

Wu, C.F.J. (1986), "Jackknife, Bootstrap, and Other Resampling Methods in Regression Analysis", *Annals of Statistics*, vol. 14, pp. 1261-1350.

Yatchew, Adonis (1998). "Nonparametric Regression Techniques in Economics", *Journal of Economic Literature*, vol. 36, pp. 669-721.

_____ (2003) Semiparametric Regression for the Applied Econometrician Cambridge University Press.

APPENDIX

A. DATA

Unlike household income and expenditure surveys conducted in other countries, the FAMEX is a recall survey. That is, the data for the FAMEX is collected in March/April of a given year but covers expenditures for the previous year. It is possible that the expenditure data used in the analysis may suffer from recall bias. In addition, data collectors make attempts to ensure that total expenditures are roughly equal to total income. In particular, income must balance expenditures to within 10% and records where expenditures exceed all sources of income by 20% or more are rejected. As a result, it is reasonable to assume that the estimates obtained for the underground economy using this method will be a lower bound estimate. The response rate for the FAMEX averages around 70%.

B. UNIT OF ANALYSIS

Ideally, the unit of analysis would be individuals as it would avoid assuming households act as single decision makers and since in Canada taxes are assessed on the individual rather than the household. In the FAMEX, however, expenditures are only surveyed at the household level and there are insufficient observations to conduct the analysis on single adult households. Additionally, as the FAMEX does not contain information regarding after tax income by income source¹, the application was conducted using gross income. After tax income is more desirable in the analysis as households are

¹ That is, the FAMEX contains information by household on total gross income and total net income but household self-employment income and other income is only available in gross terms. As household self-employment income and other income is used to calculate true household self-employment income (shown in equation (3b)), the application described in this paper could only be conducted using gross income terms.

more likely to base their expenditures on after-tax income. Further, as previously mentioned, income tax in Canada is assessed on the individual rather than on the household. As a result, households with similar gross incomes may not have comparable net income and hence may not have comparable expenditures which would lead to a biased estimate of true gross income in the first step of the approach.

The analysis was also conducted on married households living in both rural and urban areas. Limiting the analysis to households living only in urban areas resulted in insufficient observations. It is extremely likely that households in urban and rural environments have different levels of food expenditures at similar income levels for reasons that are unassociated with income under-reporting. For example, households in rural environments may be more likely to: grow food for consumption in a household garden; face reduced food prices due to the presence of local producers and suppliers; and engage in the trade of goods and services for food products. To the extent that this is true, food expenditures for rural households with no self-employment income will act as a poor counterfactual for urban households with positive self-employment income and vice versa.